社会媒体分析与智能研究进展

崔凯楠[1,2]，郑晓龙[2]，曾大军[2]，王飞跃[2]
(1. 西安交通大学电子与信息工程学院，西安，710049；
2. 中国科学院自动化研究所 复杂系统管理与控制国家重点实验室，北京，100190；)

摘要

随着互联网与移动通信技术的飞速发展，社会媒体对信息的生成、传播与消费方式产生了巨大的影响。这种深刻的变革为学术研究与技术应用带来了无限机遇与巨大挑战。近年来，社会媒体分析与智能已受到越来越多人的关注，目前正成为一个新兴且极富吸引力的研究领域。基于此，本文将主要分析并阐述社会媒体分析与智能这一领域近年来的研究工作。首先，我们将介绍观点挖掘/情感分析、社会网络分析等社会媒体信息分析工具，并描述影响力扩散、在线社区等社会媒体分析中的建模方法与预测研究。接着我们将介绍社会媒体智能的基本概念与应用情况。文章最后，我们将总结并讨论该领域所面临的若干挑战。

1 引言

互联网与移动通信技术的发展推动了社会媒体(Social Media)的崛起，新的技术平台使得内容的产生与分享能够以交互的方式进行。新鲜的应用功能，良好的用户体验和含有社会属性的服务内容加速了用户生成内容(User Generated Content, UGC)的爆发。如果从平台的角度出发，各式各样的网络应用服务勾画出社会媒体的基本功能轮廓，如表 1 所示列出了社会媒体中的主要服务类别。例如：博客、微博客、社交网络、维基百科、兴趣图谱、地理签到站点、图片/视频分享站点、社会标签站点、网络社区等等。事实上，社会媒体中产生并流动着的海量用户生成内容正是 Web2.0 时代的主要特性。在 Web2.0 时代，网民被赋予了更多的权利，他们可以发表自己的观点，或者发起某项集体行动，这使得网络中的行为愈加密集，更具风险性，同时也更为复杂。

在维基百科中，社会媒体被定义为“人们彼此之间用来分享意见、见解、经验和观点的工具和平台”[1]。广义来说，社会媒体代表了一种新的沟通方式，这是一种基于社区的分布交互式的内容产生和传播机制[2]。区别于传统媒体（例如：报纸、广播、电视）的广播式传播机制，社会媒体模糊了信息生产者与信息消费者之间的界限。这使得信息的生产、传递和消费模式发生了根本性的转变[3]。各种各样的用户生成内容（例如博客、评论、微博等）在信息的生产者和消费者的复杂网络间实时地流动着。

表 1 社会媒体主要类别

类别	国际重要站点	中国重要站点
博客	LiveJournal, WordPress	搜狐博客，新浪博客
微博客	Twitter	新浪微博，腾讯微博
商业口碑	Yelp, Epinions	大众点评
图片视频分享	Instagram, Youtube, Flickr	优酷网，土豆网，网易相册
社交网站	Facebook, Linkedin	人人网，开心网
维基百科	Wikipedia	百度百科，互动百科
地理位置服务	Foursquare	街旁
兴趣图谱	Pinterest	花瓣网，堆糖网

据统计[4]，2011 年全世界共有社会网络账户 24 亿个，Twitter 的用户数量在 2011 年已经达到 2.25 亿，每天发送的 tweet 数量为 2.5 亿条。平台中最火热的话题每秒钟可吸引 8868 条 tweet，平台最受关注的用户拥有 180 万名关注者；相比于 Twitter，Facebook 拥有更广泛的用户群体，自 2004 年创建以来，Facebook 的用户数量持续增长，如图 1 所示。Facebook 公布的最新数字显示的用户总数已经达到 9.01 亿人，每日接受到的评论为 32 亿条，每日新增图片 3 亿张。网站上的好友数量达到 1250 亿对[5]。

图 1 Facebook 用户数量增长情况（根据新闻报道重绘[6]）

2 社会媒体的机遇与挑战

当下，社会媒体中自由涌动的海量异构数据已成为一种重要的资源，它以文本、图像、多媒体等形式呈现，其潜在价值将超越以实物呈现各种资源，例如：石油、矿藏等。因此对社会媒体的挖掘与利用引发了社会各界的广泛关注，消费者、选民、商业组织、政府机构、非营利组织都可以从社会媒体中找到关联的应用方向，这使得关于社会媒体的应用和研究在短时间内得到了爆发式的增长。以公共卫生领域为例，社会媒体中实时、海量的信息可以用于感知大规模人群健康状态，其将直接服务于公共卫生事件的实时监测中，如传染病事件暴发的早期检测与预警。传统的流行病学调查方法获得的资料需要巨大的时间与人力成本，这使得实证分析的结果不能及时反映传染病的真实传播过程。而社会媒体中的任何用户都是网络信息传感器的一个细小终端，可以实时感应并报道重大疫情的发生和发展状况，这一特点可以用于弥补流行病学调查数据之不足。最引人注目的应用是 Google 公司利用搜索关键词预估流感趋势的创新性研究方法—Google Flu Trends[7]，它主要通过统计与流感症状相关的各类关键词出现频度与相关的空间信息进行预测。其预测结果与美国疾病预防控制中心的监测结果非常接近，而且实时性更强，可比传统方法提前一到两周得到监测结果。这种创新方法展现了传统手段难以企及的预测效果。Nelson 等人[8]将世界范围中与传染病相关的疫情暴发新闻通过电子地图进行实时展示，这种可视化的方式可帮助人们分析疾病暴发的时空风险。

对于政府组织而言，社会媒体在防灾应急中也展示了巨大的应用价值。最近几年，每当突发事件出现时，社会媒体中便会以极快的速度大量涌现出从各个视角对突发事件的描述、评论与应对建议。网络用户不仅将突发事件现场的各类一手资料信息通过社会媒体分享，甚至会深入挖掘出突发事件的各种相关信息。突发事件出现后，社会媒体背后各种网络群体展现出的热情与高效所产生的重大价值是传统方法所无法想象的。例如，2010 年海地和智利特大地震中，基于对 Twitter 等微博内容的内容聚合平台—Ushahidi.com[9]实现了对事件现场各类信息的即时反应、快速更新与全方面覆盖，为灾情实时感知、救援进展实时播报、搜寻幸存者等提供了具有重要价值的信息。事实上，早在 2004 年美国地质调查局(United States Geological Survey，USGS)开通了“Did You Feel It”网站，以便从网络民众直接获取地震活动相关的各类信息。2007 年，美国政府综合论坛、博客、掘客、标签、RSS、社交等社会媒体技术构建了专业政府防灾应急网“911.org”。社会媒体给非常规突发事件应急管理的创新研究提供了难得机遇。

通过追踪社会媒体中的信息，商业公司可以获得有助于推广和改进产品的反馈和信息。对于消费者，海量的信息和各种观点可以帮助他们走进群体智慧的大门，帮助他们制定更加理智的决定。社会媒体的出现极大促进了社会科学与信息科学的融合，为科学研究带来了新的机遇。社会计算采用信息与计算技术方法研究社会相关问题，例如：社会群体的形态及其演化

规律，人们的交互模式等。为了解决社会计算中复杂系统的管理与控制问题，王飞跃等人扩展了社会计算的概念[11]，并提出了面向社会计算研究的 ACP 方法[12,13]。ACP 方法通过建立人工社会，利用计算实验和平行执行分析评估影响系统的各种因素，进而实现实际系统和人工系统的相互借鉴和互动调节。这些研究可以帮助我们解答很多实际问题，如某个组织是达到了一个稳定的状态还是经常发生剧烈变化，具有创造力的团队应具有什么样的交互方式等[10]。2009 年，David Lazer 等[14]在 Science 杂志上提出计算社会学的概念，指出社会媒体中的大量信息，如博客、论坛、聊天、消费记录、电子邮件等，都是对现实社会的人及组织行为的映射，社会媒体中的数据可用来分析个人和群体的行为模式，从而深化我们对生活、组织和社会的理解。

然而社会媒体在带来巨大机遇的同时也面临着一些挑战，如海量数据的处理问题、网络隐私保护问题等。社会媒体中产生的海量数据促进了新学科的出现与发展，数据科学(Data science)是一门研究数据获取、数据存储与管理、数据安全、数据分析、可视化等内容的新兴学科。在 2008 年《Nature》杂志就出版了名为“Big Data”的专刊，从多个方面着重介绍了人们面对海量数据所遇到的困难。2011 年 1 月《Science》杂志推出专刊“Dealing with data”，同年 12 月又推出专刊“Data Replication & Reproducibility”，这两期专刊主要讨论如何应对海量数据处理中的各种问题。

不同于海量数据处理所带来的技术挑战，网络隐私保护问题需要考虑更多层面的因素。出于对日益严重的隐私保护方面问题的考虑，世界各国纷纷酝酿出台新的法律法规。2012 年 1 月 25 日，欧盟提交了新的数据保护法修订案，新数据保护法中规定泄露用户数据的公司将面临高额罚金。2012 年 2 月 23 日，美国向议会提交了新的隐私权利草案。由于法律环境的变化，一些社会媒体站点管理者开始抵制公众和第三方机构对于数据的大规模获取，例如：Foursquare(地理信息服务站点)的服务条款中明确提及“禁止自动获取站点所提供内容”，Twitter 要求斯坦福大学移除先前公布的用于研究的数据集。在数据分享方面，Fred Morstatter 等[15]总结了目前环境下社会媒体数据分享中面临的问题，并提出了一种基于开放准则的分布式数据分享框架。此外，研究人员已经开始尝试通过各种分割与合并算法降低社会网络数据公开后的隐私担忧。

3 社会媒体分析

社会媒体的研究具有天然的多学科特性，在过去的几年中，来自各主流学科的研究团体均开始关注社会媒体带来的科学机遇与挑战。如果从信息科学的角度出发，社会媒体分析和社会媒体智能是该领域中最重要的研究方向。由于社会媒体的动态性与复杂性，信息传播的全面性和快速性、网上群体行为演化的动态性、网上以及网下行为虚实交互性，对于社会媒体的分析迫切需要各种分析工具与模型的支撑。具体说来，社会媒体分析的研究包括社会媒体数据的获取与追踪方法、数据挖掘方法与分析工具、社会媒体建模方法、分析结

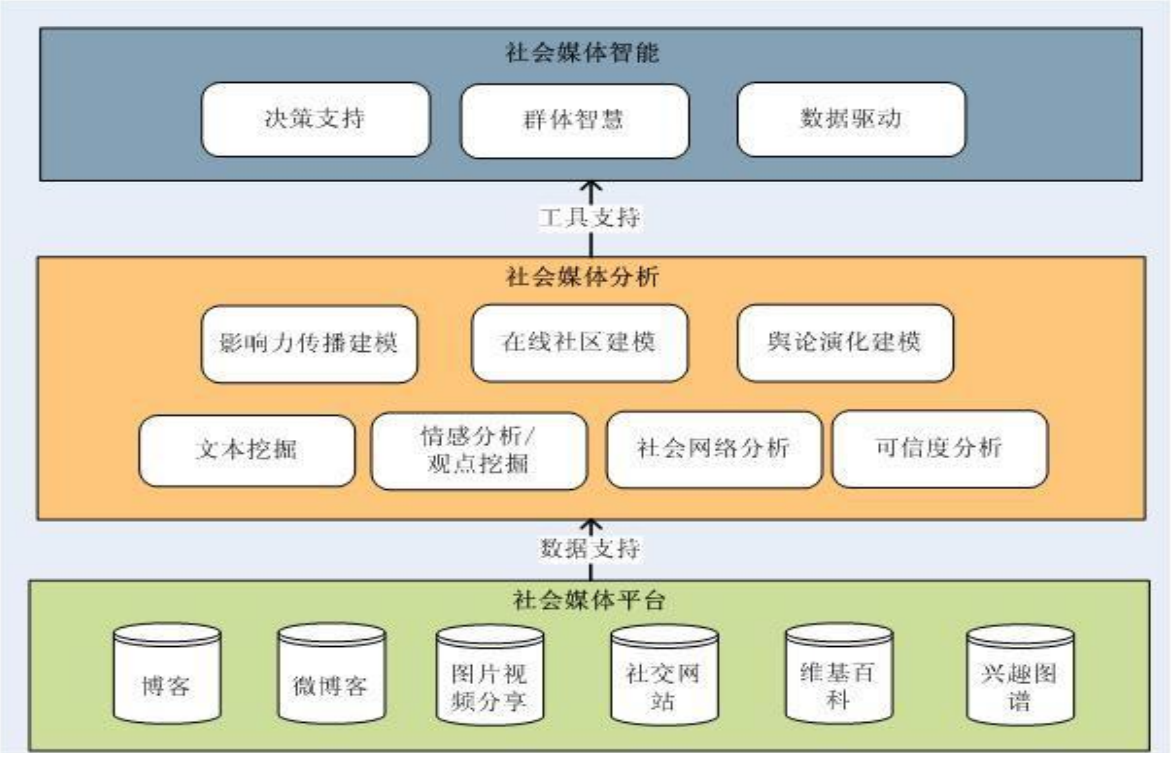

图 2 社会媒体分析与社会媒体智能

果的可视化方案等。本文将主要介绍社会媒体分析中的数据分析技术与社会媒体建模方法，如图 2 所示。

3.1 社会媒体分析的工具与方法

目前，常用社会媒体分析工具有观点挖掘[16]、社会网络分析[17]、文本挖掘、可信度分析等。随着复杂网络研究的兴起[18,19]，社会网络分析受到越来越多的关注，社会媒体中丰富的社会关系为使得开展大规模社会网络分析成为可能。此外由于社会媒体中含有大量用户生成的主观评论信息，通过情感分析和观点挖掘技术可以获取大众对于某一事件或产品的看法。情感分析/观点挖掘是对带有主观性的文本信息(以及语音识别后转化成的文本)进行分析、处理、归纳和推理的过程。简而言之就是获取意见的倾向性(正面、负面与中立)与强度信息。进行情感分析的另一个重要目的是要寻找意见和情感的主体和客体。其中情感的主体就是观点和情感的持有人或者组织。观点的客体就是观点针对的目标，例如，某事件、事件的起因、事件的影响、事件的结局等等，上述目标也可以是产品、个体、公司、组织等等。如果从粒度这一维度进行划分，可分为文档级别分析与句子级别的分析。目前进行情感分析的主要方法为基于统计/学习的方法与基于规则的方法，更多的技术是两者结合使用。

3.2 社会媒体的建模与预测

在社会媒体中，网络中的节点通过信息传播与各种交互行为直接或间接影响其它节点行为。影响力传播建模主要分析研究信息和影响力在社会媒体中的传播扩散

的规律与机制[20]。例如：Sun 等人[21]通过研究 Facebook 中的扩散现象，发现不同于“扩散是由少量初始节点发起”这一传统假设，喜爱 Facebook 中公共页面的行为通常是由大量节点同时发起，并通过很短的传播链迅速结成大集群。类似的工作中，Lerman[22]等人通过分析活跃用户分享新闻的行为，研究新闻在 Digg 和 Twitter 等站点中的传播特征。作者认为尽管两个站点在功能和界面上差异很大，但是信息传播的方式却很相似。Yang 等人[23]则通过聚类算法研究信息扩散的时序特征，通过分析 Twitter 和 blog 中的数据共得到 6 种时序曲线的主要形状，这些信息可用于进一步的预测工作。

影响力传播建模通常以某种产品、技术、话题在社会媒体中演化与发展作为研究背景，通过各种模型研究这些信息在社会媒体中的传播与扩散过程。很多研究问题均基于这些影响力传播模型，如影响力最大化、爆发检测、意见领袖识别等。Bhatt 等人[24]通过分析在线社会网络中产品购买行为的传播扩散过程，建立了相应的预测模型。该预测模型可以用于确定不同营销策略的目标人群，从而使得产品的市场份额最大化。Goyal 等人[25]通过研究 Flickr 中图片的传播行为，提出机器学习算法，通过挖掘历史传播记录获取影响力传播模型的参数。根据模型种类不同，影响力传播模型大致可分为线性阈值模型，独立级联模型与博弈演化模型三种。

独立级联模型(Independent Cascade Model)认为网络中的感染节点 i 以概率 p_{ij} 来感染其邻居节点 j，但每个节点只有一次机会来感染同一节点，直到网络中所有节点的结束传染尝试为止。

线性阈值模型认为网络中的节点均存在一个阈值 $h_i \in [0,1]$，网络中的每条边 e_{ij} 都有一个权值 w_{ij} 并且满足 $\sum_{j \in N_i} w_{ij} = 1$，其中 N_i 表示节点 i 的邻居节点。当节点 i 满足条件 $\sum_{j \in N_i,\, j \in S} w_{ji} \geq h_i$ 时被感染，其中 S 为感染节点集合。

与上述两个模型不同，博弈演化模型中并不将状态分为感染或者未感染，而使用 x_i 表示一个节点的选择，节点会根据当前状态改变自己的选择。$u(x_i, x_j)$ 表示收益函数。对于两个相连的节点而言，选择相同收益值大于选择不同的收益。网络中的每条边 e_{ij} 都有一个权值 w_{ij}，每一时刻节点 i 都会调整自己的选择使得自己的总收益 $\sum_{j \in N_i} w_{ij} u(x_i, x_j)$ 最大，其中 N_i 表示节点 i 的邻居节点。

在线社区的建模与分析主要研究群体行为的建模理论，探索在线社区的形成、演化与发展规律。例如：在线社区识别主要通过网络节点间的链接关系发现潜在的社会群体。随着信息技术的发展，恐怖组织和仇恨群体越来越多的使用社会媒体进行传播思想、招募成员等活动。为了应对这些新的挑战，Chau 等人[26]提出了一种识别博客中的仇恨团体的半自动方法，这种方法不仅可以应用于实际的信息监控中，也可以帮助研究仇恨团体的结构特征。群体行为分析的主要方法包括社会文化建模

分析和组织行为预测，此外，规划识别方法可以根据观察到的群体行为识别当前行为的目标和意图等深层信息。

4 社会媒体智能

越来越多的事实表明“群体智慧”在网络时代拥有巨大能量[27]。以 2006 年的人肉搜索事件—“网络虐猫女”为例，网友们仅凭借网络中发布的一组照片，在短短六天时间里，找到了踩猫人、光碟拍摄人等关键人物，完成了几乎不可能完成的任务。同样，作为“群体智慧”，另一种形式，众包也因其强大的影响力渗透到生活中的各个方面。亚马逊还推出了专门提供众包服务的平台 Mechanical Turk。

如果说各种人肉搜索、众包现象充分展现了“群体的智慧”的力量与魅力，社会媒体智能则是建立在社会媒体分析之上，通过整合社会媒体分析的技术、方法与模型，旨在为各种应用问题提供有效的决策支持，帮助决策者利用“群体智慧”。社会媒体智能更倾向于如何激励、引导、收集、利用这些智慧。由国防高级研究计划局（DARPA）举办的网络挑战（Network Challenge）展示了社会媒体智能的巨大潜力[28]。在该竞赛中，组织者秘密的将 10 只红色的气象气球放置到美国各地，在最短时间定位所有气球正确位置的参赛团队将获得 4 万美金的奖励。由于任务的跨地域性、时间敏感性，单单依靠个人或少数人的力量寻找所有气球几乎是不可能的。幸运的是，气球放置的位置多为各种公共场所，例如：公路旁，棒球场内等，普通民众均可以看到并提供信息。因此如何在短时间内动员人们进行大规模协作是参赛队伍面临的主要挑战。由于需要实时、动态的信息交互方式，在线资源特别是各类社会媒体站点显现出明显的优势。在社会媒体的帮助下，获胜的团队通过“递归激励”策略开展动员，利用适当的信息确认方法收集整理各种结果与信息，最终在不到 9 个小时的时间内便找到所有 10 个气象气球，这一结果远低于组织者的预计，展示了社会媒体巨大的影响力。

在对获胜策略的具体分析中，Pickard 等人展示了社会媒体在竞赛过程中发挥的重要作用[29]。通过分析 Twitter 中的数据可以发现，有些队伍主要依赖于初期密集的活动吸引参与者及注意力，但是这种方式所造成的影响力很快消失了。而获胜团队的“递归激励”策略可以使社会媒体中的影响力更加持续。该问题的实质是如何在有限的时间内开展大范围情报收集工作，因此竞赛结果将有助于思考如何开展大规模灾害，商业情报的收集。该活动也为网络社区传播理论、众包激励机制在社会媒体智能中的应用提供了良好的示例。

社会媒体分析的快速发展与广泛应用为社会媒体智能提供了坚实的基础，极大的促进了社会媒体智能在各领域中的应用。2012 年 3 月，美国红十字会正式启用了新的“数字化运营中心”，作为第一个基于社会媒体的人道主义救援部门，该中心的建成反映了社会媒体智能在应急管理中日益增长的重要性[30]。具体来说，数字化运营中心将帮助红十字会组织在紧急情况下收集各种受灾地区的信息，以确定需要救助的人们，并为他们提供更好的服务，例如帮助人们找到所需的食物、水、避难

所或是情感支持。在金融领域，2011 年 5 月，世界首家基于社交媒体的对冲基金 Derwent Capital Markets 上线，它将利用 Twitter 中的公众情绪实时分析结果指导投资[31]。这些应用的背后是相关研究的支持，目前尝试使用社会媒体预测股票市场可以分为两种，一种方法主要比较 Twitter 中的情绪信息与股票指数之间的关联性。如 Zhang[32]等发现无论是“希望”等正面情绪或是“害怕”、“担心”等负面情绪在总推文数中的比例, 都可以预示道琼斯指数、纳斯达克指数、标准普尔指数的变化。另一个研究中通过对 Twitter 进行更深入的分析，研究者发现推文的总数、交易量、看涨程度和标准普尔 100 指数之间密切相关[33]。另一种方法则比较品牌在社会媒体中的受欢迎程度与公司股价之间的关联。通过分析星巴克、可口可乐和耐克等三个品牌在社会媒体中的数据，研究人员发现公司的股价与社会媒体中的社会网络信息密切相关，这些指标包括 Facebook 上的粉丝数、Twitter 上的听众数和 Youtube 上的观看人数等[34]。

然而，与社会媒体分析相比，社会媒体智能的研究仍然处于起始阶段，尽管目前存在着许多概念层面的讨论，例如相关的技术挑战，相关的学科等，然而专注于社会媒体智能的系统性研究与细致而可信的研究结论依然非常稀少。在这里我们尝试讨论社会媒体智能的研究面临的主要挑战，以期该领域的研究能够得到更广泛的关注。首先，社会媒体智能的研究迫切需要不同学科之间的交互与整合。尽管以商业为代表的各种领域已经对社会媒体智能的研究表现出极大的兴趣，但是学科之间的整合与交流程度仍然很低，信息科学的研究方法与研究问题仍然占据着最重要的位置。相较而言，社会媒体智能在心理学、社会心理学、传媒学、政治学等领域的研究比较分散，缺乏相应的跨学科项目支持。其次，对于社会媒体智能的研究需要制定客观和简明的性能评价方法。由于社会媒体智能的应用领域广泛，建立一套有效的决策支持评价机制是非常困难的。这种方法上的困难使得我们很难评价社会媒体智能的投入产出比，因而阻碍着相关研究的开展。最后，如果单纯从模型建立和决策支持的角度出发，随着社会媒体智能研究的不断成熟与发展，新的应用也将需要更先进的社会媒体分析方法与计算框架的支持。

5 总结

综上所述，在信息科学的研究中，社会媒体的分析与智能引起了广泛的关注并取得了丰硕的成果。社会媒体分析主要关注社会媒体数据的跟踪与获取、社会媒体数据的处理与整合、社会媒体的建模与预测等问题。在社会媒体分析的研究中，情感分析/观点挖掘、社会网络分析等方法与技术的成熟推动着社会媒体建模与预测的研究，诸如影响力传播建模、在线社区建模等研究均取得了一定的进展。社会媒体智能的研究建立在社会媒体分析之上，通过整合社会媒体分析的技术、方法和模型，为各种应用问题提供有效的决策支持，帮助决策者利用“群体的智慧”。与社会媒体智能的巨大应用价值相比，其研究仍然处于早期阶段，因此需要广泛的关注与深入的讨论。

致谢

本文作者感谢以下项目支持，国家自然科学基金委(NNSFC)项目：71103180, 91124001, 60921061, 90924302, 70890084, 91024030 and 40901219; 以及中国科学院项目(CAS)2F11N06。

参考文献

[1] 社会媒体-维基百科. Available: http://zh.wikipedia.org/wiki/%E7%A4%BE%E4%BC%9A%E5%AA%92%E4%BD%93

[2] D. Zeng, et al., "Social media analytics and intelligence," Intelligent Systems, IEEE, vol. 25, pp. 13-16, 2010.

[3] J. Leskovec, "Social media analytics," presented at the Proceedings of the 17th ACM SIGKDD International Conference Tutorials, San Diego, California, 2011.

[4] Internet in Numbers 2011. Available: http://royal.pingdom.com/2012/01/17/internet-2011-in-numbers/

[5] Facebook passes 900 million users. Available: http://www.techweekeurope.co.uk/news/facebook-passes-900-million-users-74646

[6] 增长:Facebook 最惊人的力量. Available: http://www.36kr.com/p/80741.html

[7] J. Ginsberg, et al., "Detecting influenza epidemics using search engine query data," Nature, vol. 457, pp. 1012-1014, 2009.

[8] R. Nelson, "HealthMap: the future of infectious diseases surveillance," Lancet Infectious Diseases, vol. 8, pp. 596-596, 2008.

[9] O. Okolloh, "Ushahidi, or'testimony': Web 2.0 tools for crowdsourcing crisis information," Participatory Learning and Action, vol. 59, pp. 65-70, 2009.

[10] X. Zheng, et al., "Next-Generation Team-Science Platform for Scientific Collaboration," Intelligent Systems, IEEE, vol. 26, pp. 72-76, 2011.

[11] F.-Y. Wang, et al., "Social Computing: From Social Informatics to Social Intelligence," Intelligent Systems, IEEE, vol. 22, pp. 79-83, 2007.

[12] F.-Y. Wang, "Toward a Paradigm Shift in Social Computing: The ACP Approach," Intelligent Systems, IEEE, vol. 22, pp. 65-67, 2007.

[13] F.-Y. Wang, "Computational experiments for behavior analysis and decision evaluation of complex systems," Journal of system simulation, vol. 16, p. 5, 2004.

[14] D. Lazer, et al., "Life in the network: the coming age of computational social science," Science (New York, NY), vol. 323, p. 721, 2009.

[15] F. Morstatter, et al., "Opening Doors to Sharing Social Media Data," IEEE Intelligent Systems, pp. 47-51, 2012.

[16] B. Pang and L. Lee, "Opinion mining and sentiment analysis," Foundations

and Trends in Information Retrieval, vol. 2, pp. 1-135, 2008.

[17] P. J. Carrington, et al., Models and methods in social network analysis: Cambridge Univ Pr, 2005.

[18] A.-L. Barabási and R. Albert, "Emergence of Scaling in Random Networks," Science, vol. 286, pp. 509-512, October 15, 1999 1999.

[19] X. Zheng, et al., "Analyzing open-source software systems as complex networks," Physica A: Statistical Mechanics and its Applications, vol. 387, pp. 6190-6200, 2008.

[20] D. Kempe, et al., "Maximizing the spread of influence through a social network," presented at the Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, Washington, D.C., 2003.

[21] E. Sun, et al., "Gesundheit! modeling contagion through facebook news feed," 2009.

[22] K. Lerman and R. Ghosh, Information Contagion: An Empirical Study of the Spread of News on Digg and Twitter Social Networks, 2010.

[23] J. Yang and J. Leskovec, "Patterns of temporal variation in online media," presented at the Proceedings of the fourth ACM international conference on Web search and data mining, Hong Kong, China, 2011.

[24] R. Bhatt, et al., "Predicting product adoption in large-scale social networks," presented at the Proceedings of the 19th ACM international conference on Information and knowledge management, Toronto, ON, Canada, 2010.

[25] A. Goyal, et al., "Learning influence probabilities in social networks," presented at the Proceedings of the third ACM international conference on Web search and data mining, New York, New York, USA, 2010.

[26] M. Chau and J. Xu, "Mining communities and their relationships in blogs: A study of online hate groups," International Journal of Human-Computer Studies, vol. 65, pp. 57-70, 2007.

[27] F.-Y. Wang, et al., "A Study of the Human Flesh Search Engine: Crowd-Powered Expansion of Online Knowledge," Computer, vol. 43, pp. 45-53, 2010.

[28] DARPA Network Challenge. Available: http://en.wikipedia.org/wiki/DARPA_Network_Challenge

[29] G. Pickard, et al., "Time-critical social mobilization," Science, vol. 334, pp. 509-512, 2011.

[30] Social Media Digital Operations Center for Humanitarian Relief. Available: http://content.dell.com/hk/en/corp/d/corp-comm/red-cross-digital-operations

[31] The World's First Twitter-Based Hedge Fund Is Finally Open for Business. Available: http://www.theatlantic.com/business/archive/2011/05/the-worlds-first-twitter-based-

hedge-fund-is-finally-open-for-business/239097/

[32] X. Zhang, et al., "Predicting Stock Market Indicators Through Twitter “I hope it is not as bad as I fear”," Procedia-Social and Behavioral Sciences, vol. 26, pp. 55-62, 2011.

[33] T. Sprenger and I. Welpe, "Tweets and trades: The information content of stock microblogs," 2010.

[34] Study Claims Stock Price, Social Media Are Linked. Available: http://www.pcmag.com/article2/0,2817,2382391,00.asp

作者简介

崔凯楠：西安交通大学电信学院在读博士研究生。2007 年获西安交通大学学士学位，2009 起赴中国科学院自动化研究所复杂系统管理与控制国家重点实验室进行联合培养至今。主要研究方向为社会媒体分析。

郑晓龙：中国科学院自动化所研究所助理研究员。2003 年获得中国计量学院学士学位，2006 年获得北京交通大学硕士学位，2009 年获得中国科学院博士学位，2009 年起，在中国科学院自动化所复杂系统管理与控制国家重点实验室工作至今。主要从事社会网络与数据挖掘方面的研究与相关技术研发工作。是 IEEE、ACM Member 以及，Secretary of ACM social and economic chapter，获得国家省部级奖项多项。

曾大军：中国旅美科技协会总会 2012 年理事会主席，2010 年度会长,中国科学院自动化研究所研究员、博士生导师及 Web 科学及服务计算实验室主任。1990 年毕业于中国科学技术大学少年班、经济管理与系统科学系及计算机系，获学士学位；1998 年获美国卡内基梅隆大学工业管理博士学位。他主要研究领域为时空数据分析、传染病信息学、情报与安全信息学、多智能体系统及社会计算。

王飞跃：中国旅美科技协会总会 2005 年度会长，中国科学院自动化研究所研究员、博士生导师及复杂系统管理与控制国家重点实验室主任。1982 年获山东化工学院学士学位，1984 年获浙江大学硕士学位，1990 年获美国伦塞利尔理工学院计算机与系统工程博士学位。他主要研究领域为智能系统和复杂系统的建模、分析和控制，是智能控制方面的国际知名学者，也是该领域的早期开拓者之一。